

MIDI音源を用いた旋律の言語依存性に関する分析

三好 真人^{1,a)} 柘植 覚^{2,b)} 武田 一哉^{1,c)}

概要：本稿では、作詞者と作曲者の国籍および出身が日本と英語圏であるポピュラー音楽の MIDI 音源を用いて、旋律における言語的特徴を捉えるための音高遷移モデルを提案する。提案手法では、MIDI 音源から得られる音程系列と音名系列に対し、統計的言語モデルである N -gram モデルと Nested Pitman-Yor 言語モデルを用いて音高遷移をモデル化する。提案手法の有効性を検証するために、日本語と英語ポピュラー音楽の MIDI 音源 212 曲を用いて、各言語での音高遷移の傾向調査ならびに言語の自動識別実験を行った。調査結果より、日本語と英語で異なる音高遷移特徴が存在することが示唆された。また識別実験より、提案手法による平均識別率は 81.2%であることが分かった。

1. はじめに

現在、CD、音楽配信システムなどにより膨大な音楽が容易に入手可能である。膨大な音楽の中から利用者が所望する音楽を見つけるために有効な音楽検索システムが求められている。一般的に音楽検索システムのクエリとしては、楽曲のタイトルやアーティスト名など音楽に関する情報が用いられる。さらに有効で柔軟な音楽検索システムを実現するためには、音楽内容（音色、歌詞、楽器構成など）に関する情報をクエリとすることが必要であると考えられる。そのために、音楽内容の正確なモデル化が必要である。

本研究では、作詞者と作曲者の母国語が旋律に影響することにより、言語に特有な旋律が構成されると仮説する。ここで言語に特有な旋律とは日本語らしい旋律や英語らしい旋律を示し、このような旋律における言語的特徴を抽出することを目的とする。作詞者と作曲者の母国語のどちらが旋律に強く影響を与えるかについては明確ではないため、本稿では作詞者と作曲者が共に同じ母国語の場合のみを扱う。音楽から言語的特徴が抽出することができれば、音楽に言語情報を与えることが可能となり、音楽検索システムへの応用が期待できる。言語と音楽には深い関連性が存在することが指摘されている。言語には文法、音楽には音楽理論という一定の構築規則が共通して存在することが文献 [1], [2] で述べられている。また、言語のアクセントと

旋律の関連性を調査した研究 [3], [4] や複数の言語間で旋律特徴の違いを調査した研究 [5], [6], [7], [8], [9], [10], [11] が報告されている。さらに、複数の言語間で旋律の音程特徴を自動抽出する試みも報告されている [12]。これら先行研究では、二つの音符間の音高や音価変化の特徴や旋律における強拍位置と音声言語における強勢アクセント位置の関連性を調査することが主としてなされており、音楽検索への応用を目的とした統計的確率モデルによる旋律のモデル化はなされていない。また統計的確率モデルを用いた研究として、可変長マルコフモデルを用いた日本民謡に特有な音程系列の抽出 [12] があるが、ポピュラー音楽への適用ならびに多言語識別への有効性についての検証は行われていない。一般的に聴取されるポピュラー音楽を対象に旋律の言語的特徴抽出の検証、ならびに多言語識別の検証は、音楽検索への応用のために重要であると考えられる。

本稿では、作詞者と作曲者の母国語が旋律の音高遷移に影響すると仮説し、旋律を統計的言語モデルである N -gram モデルと Nested Pitman-Yor 言語モデルによりモデル化することを提案する。音高遷移として、音符間の音高変化を捉えるための音程遷移と音階を考慮した音高変化を捉えるための音名遷移を使用する。日本語と英語のポピュラー音楽の MIDI 音源を用い、各言語における旋律の音程遷移および音名遷移の統計的確率モデルを学習し、言語による旋律特徴を調査する。また、学習した統計的確率モデルを用いた言語識別実験を行い、提案手法の言語識別性能を検証する。

¹ 名古屋大学大学院
Nagoya University, Nagoya, Aichi 464-8603, Japan

² 大同大学
Daido University, Nagoya, Aichi 457-8530, Japan

a) miyoshi.masato@g.sp.m.is.nagoya-u.ac.jp

b) tsuge@daido-it.ac.jp

c) kazuya.takeda@nagoya-u.jp

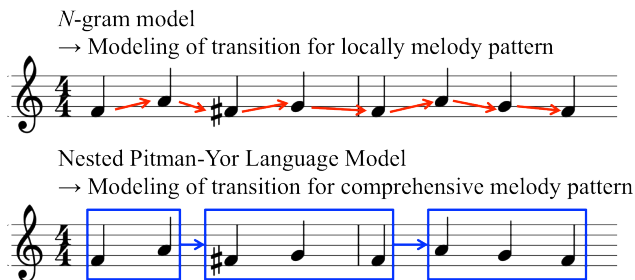


図 1 統計的言語モデルによる音程遷移のモデル化

2. 統計的言語モデルを用いた音高遷移のモデル化

本節では、統計的言語モデルである N -gram モデルと Nested Pitman-Yor 言語モデル (NPYLM) を用いた音高遷移のモデル化について説明する。本稿では、音高遷移として音符間の音高変化を捉えるための音程遷移と音階を考慮した音高変化を捉えるための音名遷移を使用する。 N -gram モデルによる音高遷移モデルは局所的な音高遷移を表現し、NPYLM による音高遷移モデルは複数音符を一つのパターンとして捉え、このパターンの遷移をモデル化し、大局的な音高遷移を表現する。次節以降で各手法についての詳細を述べる。

2.1 MIDI 音源からの音程系列と音名系列の抽出

2.1.1 音程系列の抽出

MIDI 音源の多くは Standard MIDI Format (SMF) であり、この形式は MIDI イベントとそのイベントが継続する時間 (デルタタイム) が一組として表現されている。本稿では、SMF 中のイベントのノート番号と発音のタイミングを音高系列として使用する。また、無音区間が 8 分音符以上の長さであった場合を休符とした。本稿では、図 2 に示す通り休符により区切られる区間を一つの区間として音程を抽出する。区間 l ($l = 1, \dots, L$) における j 番目のノート番号を $n_{l,j}$ とすれば、 i 番目の音程 d_i は次式で計算される。

$$d_i = n_{l,j+1} - n_{l,j} \quad (1 \leq l \leq L) \quad (1)$$

$$i = \sum_{k=1}^{l-1} G_P(k) + j \quad (1 \leq j \leq G_P(l)) \quad (2)$$

ここで、 L は総区間数、 $G_P(k)$ は k 番目の区間における音程数を示す。本稿では 2 オクターブに相当する音程の範囲を考慮するため、集合 $S_I = \{s \in \mathbb{Z} | -24 \leq s \leq 24\}$ を用いて、 $d_i \in S_I$ である。例えば、 $d_i = 0$ であれば完全 1 度 (同じ音高) であることを示し、 $d_i = 5$ であれば完全 4 度 (全音 2 個と半音 1 個) であることを示す。

2.1.2 音名系列の抽出

本稿では、旋律における調は全て既知であると仮定する。2.1.1 節と同様にノート番号の系列を音高系列として使用する。MIDI 音源に含まれる調情報を用いて、旋律の調を

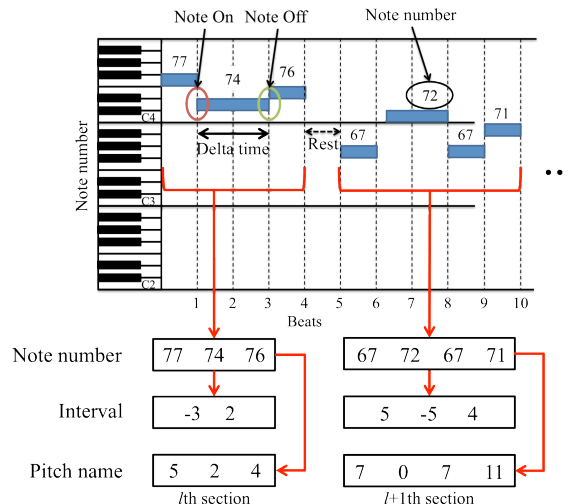


図 2 MIDI 音源からの音程系列と音名系列の抽出

全て C に移調する。ここでは、区間 l における移調後のノート番号を $\tilde{n}_{l,j}$ とする。次にオクターブを無視するようにノート番号を変換する。これは次式で表現される。

$$x_i = \tilde{n}_{l,j} \bmod 12 \quad (1 \leq l \leq L) \quad (3)$$

$$i = \sum_{k=1}^{l-1} G_P(k) + j \quad (1 \leq j \leq G_P(l)) \quad (4)$$

ここで、 $G_P(k)$ は k 番目の区間における音名数を示す。集合 $S_P = \{s \in \mathbb{Z} | 0 \leq s \leq 11\}$ を用いて、 $x_i \in S_P$ である。

2.2 N -gram モデルと Nested Pitman-Yor 言語モデルを用いた音高遷移のモデル化

上述した音程系列と音名系列を用いて、統計的言語モデルである N -gram モデルと Nested Pitman-Yor 言語モデルによりモデル化する。以降では音程系列を用いた場合のみについて述べるが、音名系列の場合は数式中の d_i を x_i 、集合 S_I を S_P に変更する。

2.2.1 N -gram モデルによる音高遷移のモデル化

MIDI 音源から抽出された音程系列 d_i を N -gram モデルによりモデル化する。具体的には、連続する N 個の音程 $d_i \dots d_{i+N-1}$ ($d_i \in S_I$) が出現する頻度をカウントし、各音程系列の相対頻度を求める。これは次式で表現される。

$$p(d_i \dots d_{i+N-1}) = (1 - \lambda) \frac{C(d_i \dots d_{i+N-1})}{\sum_{d_i \in S_I} \dots \sum_{d_{i+N-1} \in S_I} C(d_i \dots d_{i+N-1})} + \lambda \frac{1}{|S_I|^N} \quad (5)$$

ここで、 $C(d_i \dots d_{i+N-1})$ は音程系列 $d_i \dots d_{i+N-1}$ が出現した回数を示す。 λ はフロアリング定数であり、 $\lambda = 10^{-5}$ である。

2.2.2 Nested Pitman-Yor 言語モデルを用いた音程遷移モデル

Nested Pitman-Yor 言語モデル (NPYLM) は、文字系列

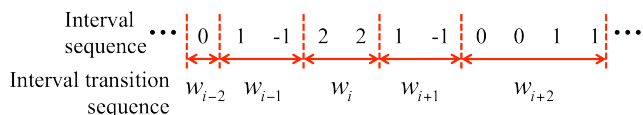


図 3 音程系列に対する音程遷移系列への分割

に対する単語分割を行うために提案された言語モデル [13] であり、文字系列に対する単語分割と言語モデルの学習を繰り返すことにより得られる。NPYLM は単語 N -gram モデルと文字 N -gram モデルから構成される。単語 N -gram モデルと文字 N -gram モデルに階層的 Pitman-Yor 言語モデル [14] を用いて、単語 N -gram モデルの基底測度を文字 N -gram モデルとすることで、文字系列から未観測な単語を推定する。NPYLM では学習時に、ギブスサンプリングに基づき確率的に単語分割位置のサンプリングを行うことで各文章に対して単語分割を行う。その後、単語分割された文章を用いて、再度 NPYLM (単語 N -gram モデルと文字 N -gram モデル) を学習する。この NPYLM の学習と単語分割を繰り返し、与えられた文章集合の生成確率が最大となる単語 N -gram モデルと単語分割を得る。本稿では、文字系列の代わりに、2.1.1 節で述べた音程系列を入力として、旋律における単語 N -gram モデルを学習する。学習された単語 N -gram モデルを用いて、ピタビアルゴリズムにより音程系列のセグメンテーションを行う。これにより、図 3 に示されるような一つの音程系列が可変長な系列長を持つ音程遷移へとセグメンテーションされる。 i 番目のセグメンテーションされた音程遷移 w_i は次式で表現される。

$$w_i = d_m \cdots d_{m+M_i-1} \quad (6)$$

$$m = \sum_{k=1}^{i-1} M_k + 1 \quad (7)$$

ここで M_k は k 番目の音程遷移における系列長である。2.2.1 節で述べた N -gram モデルと同様に、音程遷移の系列 $w_i \cdots w_{i+N-1}$ をモデル化する。音程遷移の集合を W_1 とすれば、次式で表現される。

$$p(w_i \cdots w_{i+N-1}) = (1 - \lambda) \frac{C(w_i \cdots w_{i+N-1})}{\sum_{w_i \in W_1} \cdots \sum_{w_{i+N-1} \in W_1} C(w_i \cdots w_{i+N-1})} + \lambda \frac{1}{|W_1|^N} \quad (8)$$

3. 音高遷移モデルを用いた言語識別手法

本節では、2 節で述べた音高遷移モデルを用いた言語識別手法について述べる。提案手法は識別方法により異なり、次の 3 つを提案する。

- (1) 最近傍法に基づく識別手法 (3.1 節)
- (2) サポートベクターマシンに基づく識別手法 (3.2 節)
- (3) 対数尤度に基づく識別手法 (3.3 節)

次節以降で各手法についての詳細を述べる。以降では音程遷移モデルについてのみ述べるが、音名遷移モデルについても同様の枠組みで識別する。

3.1 対称化 Kullback-Leibler ダイバージェンスを用いた最近傍法による識別

本識別手法では、まず式 (5) により各言語の音楽データを用いて音程遷移モデルを計算する。次に未知の音楽データについても同様に式 (5) により音程遷移モデルを計算し、各言語の音程遷移モデルとの距離計算により識別する。距離尺度として、本稿では対称化 Kullback-Leibler ダイバージェンスとした。各言語 ($l = 1, 2, \dots, L$) の音程遷移モデル $p_l(d_1 \cdots d_N)$ と未知の音楽データの音程遷移モデル $p_U(d_1 \cdots d_N)$ の対称化 Kullback-Leibler ダイバージェンスは式 (9) で計算される。

$$KL(p_l || p_U) = \sum_{d_1 \in S_1} \cdots \sum_{d_N \in S_1} p_l(d_1 \cdots d_N) \log \left(\frac{p_l(d_1 \cdots d_N)}{p_U(d_1 \cdots d_N)} \right) + \sum_{d_1 \in S_1} \cdots \sum_{d_N \in S_1} p_U(d_1 \cdots d_N) \log \left(\frac{p_U(d_1 \cdots d_N)}{p_l(d_1 \cdots d_N)} \right) \quad (9)$$

最も距離が近い言語を識別結果 l^* とする。

$$l^* = \operatorname{argmin}_{l=1,2,\dots,L} \{KL(p_l || p_U)\} \quad (10)$$

NPYLM における音程遷移モデルについても同様の枠組みで識別する。

3.2 サポートベクターマシンによる識別

本識別手法では、式 (5) により各音楽データについて音程遷移モデルを計算し、音程遷移モデルの確率値を特徴量として使用することでサポートベクターマシン (Support Vector Machine: SVM) を学習する。未知の音楽データに対しても同様に特徴量を得て、学習された SVM により識別する。特徴ベクトルの次元数は許容する音程の範囲 (音程集合の要素数) $|S_1|$ に対する N -gram 長のべき乗 $|S_1|^N$ となる。NPYLM の場合は、音程遷移集合の要素数 $|W_1|$ に対する N -gram 長のべき乗 $|W_1|^N$ となる。

3.3 言語モデルの対数尤度による識別

本識別手法では、各言語の音楽データを用いて音程系列の条件付き確率 $p(d_i | d_{i-N+1}, \dots, d_{i-1})$ を計算し、未知の音楽データから得られた音程系列 d_1, \dots, d_I に対する N -gram 確率を次式で計算することにより識別する。

$$p(d_1, \dots, d_I) = \prod_i^I p(d_i | d_{i-N+1}, \dots, d_{i-1}) \quad (11)$$

ここで I は音程系列長である。最も対数尤度が高い言語を識別結果 l^* とする。

$$l^* = \operatorname{argmax}_{l=1,2,\dots,L} \{\log(p(d_1, \dots, d_I))\} \quad (12)$$

NPYLM の場合は単語 N -gram 確率を用いて識別するため、音程系列についてセグメンテーション可能なあらゆるパターンについて周辺化することにより対数尤度を計算し、式 (12) と同様に最大の対数尤度を与える言語を識別結果とする。

4. 評価実験

4.1 実験条件

本実験では、国籍または出身が日本または英語圏の国（アメリカ合衆国、カナダ、イングランド、スコットランド、アイルランド）である作曲家および作詞家が創作したポピュラー音楽の市販 MIDI 音源を使用する。以降では日本の作詞家および作曲家のポピュラー音楽を日本語ポピュラー音楽、英語圏の作詞家および作曲家のポピュラー音楽を英語ポピュラー音楽と呼ぶ。音楽データ数は日本語ポピュラー音楽は 110 曲、英語ポピュラー音楽は 102 曲の合計 212 曲である。表 1 に各年代における音楽データ数を示す。各言語のポピュラー音楽に含まれるアーティストの数は、日本語が 34、英語が 41 である。アーティスト 1 組につき 1 曲から 5 曲の音楽データを含んでいる。

式 (5) における N -gram 長は 1, 2, 3 に設定した。式 (8) における N -gram 長は 1, 2, NPYLM における単語 N -gram 長と文字 N -gram 長は 3 に設定した。 N -gram モデルの対数尤度による識別では、 N -gram 長を 1, 2, 3, 4, 5, に設定し、平滑化手法として Witten-Bell 平滑化を用いた。 N -gram 確率計算の実装には SRI Language Modeling Toolkit (SRILM) [15] を使用した。NPYLM の対数尤度による識別では、単語 N -gram 長を 1, 2, 3, 4, 5, 文字 N -gram 長を 3 に設定した。NPYLM を用いた対数尤度による識別では、単語 N -gram 長は 1, 2, 3, 4, 5 に設定した。ギブスサンプリングの繰り返し回数は 200 回に設定し、NPYLM の実装には latticelm[16] を使用した。本実験における SVM は LIBSVM[17] により実装した。SVM のカーネルには線形カーネルおよび多項式カーネルを使用し、ソフトマージンパラメータはともに 10^0 から 10^4 まで 10 倍刻みで設定した。多項式カーネルの次数は 2 次、定数項は 0 に設定した。

まず、全ての音楽データを用いた各言語の音程遷移と音名遷移を比較するためのクローズ実験を行う。クローズ実験では、式 (5) により得られる各音高遷移モデルの確率値を比較する。また、各言語について NPYLM を学習した場合に学習過程で得られる音高遷移を比較する。次に、オープン実験（自動言語識別実験）を行う。オープン実験は 10-fold Cross Validation により行った。このとき、評価データとなる音楽データのアーティストが訓練データとなる音楽データのアーティストに含まれないように実験を行った。

4.2 クローズ実験結果

4.2.1 各言語における音程遷移モデルと音名遷移モデルの比較

図 4 に各言語におけるバイグラム音程遷移モデルを示す。図 4 の左図と中央図は日本語と英語における音程遷移モデルを示し、右図は日本語と英語の音程遷移モデルにおける対数確率の差を示す。図 4 の横軸と縦軸は音程を示す。数値は対数確率である。日本語と英語の音程遷移モデルを比較すると、日本語および英語ともに中心部分是对数確率が高い傾向が見られる。また、右図より、僅かではあるが中心部分より外側で日本語の対数確率が高い傾向が見られる。これより、日本語の方が音程変化の幅が大きく、英語の方が音程変化の幅が小さいということが分かった。

図 5 に各言語におけるバイグラム音名遷移モデルを示す。図 5 の左図と中央図は日本語と英語における音名遷移モデルを示し、右図は日本語と英語の音名遷移モデルにおける対数確率の差を示す。図 5 の横軸と縦軸は、遷移元の音名と遷移先の音名を示す。数値は対数確率である。日本語と英語の音名遷移モデルを比較すると、比較的類似する傾向が見られる。一方、右図より、各言語において確率が高い音名遷移が存在することも分かる。例えば、A から G#へへの遷移は日本語の方が高い確率を持つことが分かる。

4.2.2 各言語の NPYLM で獲得された音高遷移

各言語の NPYLM の学習過程で得られた音高遷移を比較する。ここでは次節で述べるオープン実験において、音程遷移で最も高い識別精度が得られた単語 N -gram 長が 3 であるときの場合について述べる。音高遷移の総数は、音程遷移では日本語で 25,901、英語で 21,439、音名遷移では日本語で 27,387、英語で 23,171 であった。また、系列長が 1 または 2 の音程遷移が日本語で 80.6%、英語で 89.4%、音名遷移が日本語で 75.0%、英語で 87.7% 全体に対して占める結果となった。表 2 と表 3 に各言語における出現頻度が多かった音高遷移を示す。ただし、系列長が 3 以上の音高遷移を対象とした。表 2 と表 3 における確率値は音高遷移の総数に対する相対頻度である。表 2 より、日本語と英語ともに同じ音程が継続する音程遷移が最も多く出現することが分かる。また表 3 より、英語では同じ音名が継続する音名遷移が多く出現しやすいことが分かる。

4.3 オープン実験結果

表 4 と表 5 に各識別手法における平均識別率を示す。表 4 と表 5 において、“Nearest Neighbor” は最近傍法による識別、“Linear SVM” と “Polynomial SVM” は SVM による識別、“Likelihood” は対数尤度による識別を示す。表 4 より、 N -gram モデルの対数尤度による識別手法が最も高い識別精度を与えることが分かった。 N -gram モデルと NPYLM を比較すると、 N -gram モデルに基づく識別手法の方が高い識別精度を与えることが分かった。これより、

表 1 各年代における音楽データ数

年代	1960年代	1970年代	1980年代	1990年代	2000年代	2010年代	合計
日本語	0	3	9	41	47	10	110
英語	10	28	21	12	23	8	102

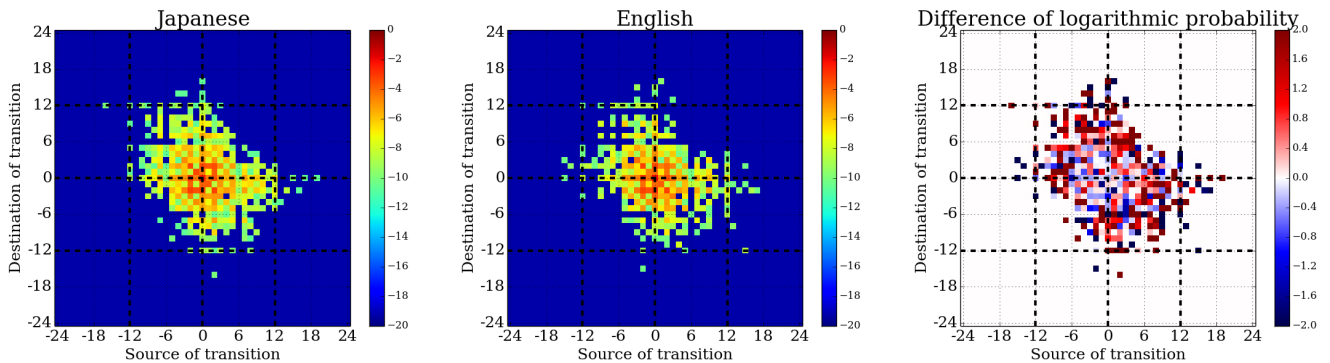


図 4 日本語と英語の MIDI 音源におけるバイグラム音程遷移モデル

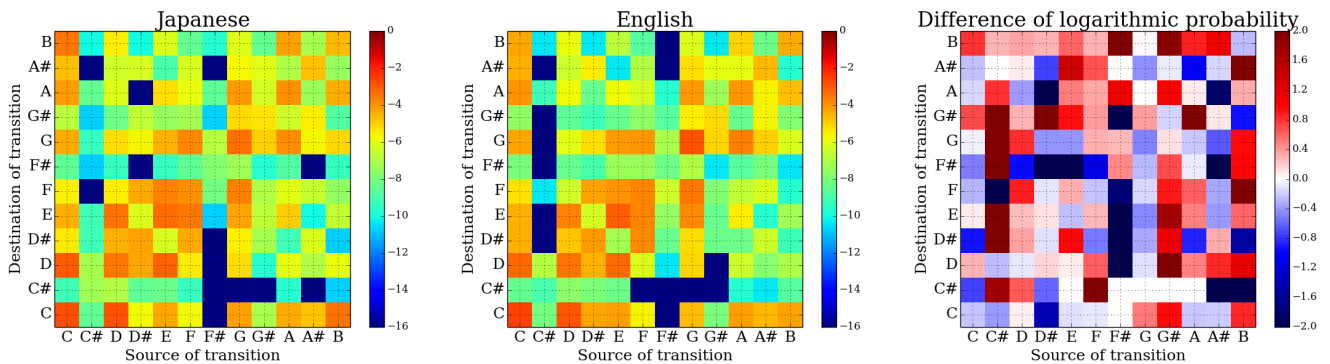


図 5 日本語と英語の MIDI 音源におけるバイグラム音名遷移モデル

表 2 NPYLM で獲得された音程遷移の確率値上位 5 位
 までの音程遷移 (系列長が 3 以上の音程遷移のみ対象)

日本語		英語	
音程遷移	確率値	音程遷移	確率値
0 → 0 → 0	0.009189	0 → 0 → 0	0.008163
0 → -2 → 0	0.007451	0 → 0 → 2	0.005224
0 → 0 → 0 → 0	0.005328	1 → -1 → -2	0.002939
1 → -1 → -2	0.005174	-2 → 0 → -2	0.002939
2 → -2 → -2	0.004787	-2 → -1 → -2	0.002519

表 3 NPYLM で獲得された音名遷移の確率値上位 5 位
 までの音名遷移 (系列長が 3 以上の音名遷移のみ対象)

日本語		英語	
音名遷移	確率値	音名遷移	確率値
A → B → C	0.008398	G → G → G	0.008329
C → D → E	0.007704	C → C → C	0.005006
C → C → C	0.007485	D# → D → C	0.004273
C → B → C	0.006280	E → E → E	0.003798
E → F → G	0.006061	F → E → D	0.003280

局所的な音程遷移が各言語の旋律特徴を特徴付ける要素であると考えられる。また、NPYLM において N -gram 長が 3 以上となると特徴ベクトルの次元数が非常に高次元となるため計算量の観点から実験ができなかった (表中の

“—”)。表 5 より、多項式カーネル SVM が最も良い識別精度を与えることが分かった。また、音程遷移と同様に、 N -gram モデルと NPYLM を比較すると、 N -gram モデルに基づく識別手法の方が高い識別精度を与えることが分かった。音程遷移と音名遷移を比較すると、音程遷移に基づく識別手法の方が高い識別精度を与えることが分かった。これより、音符間の音高変化のパターンが日本語と英語の旋律を特徴付ける要素であると考えられる。

5. おわりに

本稿では、作詞者と作曲者の母国語が旋律に影響することを調査するため、統計的言語モデルである N -gram モデルと NPYLM による音高遷移のモデル化手法を提案した。提案手法を評価するために、作詞者と作曲者の国籍および出身が日本と英語圏であるポピュラー音楽の MIDI 音源を用いて、音高遷移の調査と自動識別実験を行った。調査結果より、各言語に異なる音高遷移の傾向があることを示唆する結果が得られた。また識別実験結果より、 N -gram モデルの対数尤度による識別手法で 81.2% の平均識別率が得られることが分かった。

今後は音価の遷移についても考慮することが必要であると考えられ、音価を含めた旋律特徴について検討する。

表 4 音程遷移における識別実験結果 (平均識別率 [%])

N-gram 長	N-gram + Nearest Neighbor	N-gram + Linear SVM	N-gram + Polynomial SVM	NPYLM + Nearest Neighbor	NPYLM + Linear SVM	NPYLM + Polynomial SVM
1	68.9%	76.2%	77.8%	54.8%	74.5%	75.2%
2	70.8%	79.8%	80.2%	73.8%	69.3%	67.6%
3	73.6%	78.6%	74.7%	—	—	—

N-gram 長	N-gram Likelihood	NPYLM Likelihood
1	75.5%	74.4%
2	78.3%	69.8%
3	81.2%	76.0%
4	78.3%	73.2%
5	71.8%	75.9%

表 5 音名遷移における識別実験結果 (平均識別率 [%])

N-gram 長	N-gram + Nearest Neighbor	N-gram + Linear SVM	N-gram + Polynomial SVM	NPYLM + Nearest Neighbor	NPYLM + Linear SVM	NPYLM + Polynomial SVM
1	57.9%	69.0%	67.8%	60.8%	67.8%	71.6%
2	67.2%	72.8%	74.9%	65.8%	70.3%	68.4%
3	74.3%	74.3%	71.7%	—	—	—

N-gram 長	N-gram Likelihood	NPYLM Likelihood
1	61.1%	72.8%
2	70.2%	68.3%
3	74.6%	71.3%
4	73.5%	71.9%
5	74.5%	68.5%

謝辞 本研究の一部は、JST-CREST の支援により行われた。

参考文献

[1] 東条敏：“音楽と言語の構造認知”，情報処理学会誌，Vol. 49, No. 9, pp. 1099–1105 (2008).

[2] PhilipBall (夏目大 訳)：“音楽の科学”，河出書房新社 (2011).

[3] Nichols, E., Morris, D., Basu, S. and Raphael, C.: Relationships Between Lyrics and Melody in Popular Music, *Proc. of 10th International Society for Music Information Retrieval Conference*, pp. 471–476 (2009).

[4] Chen, X. and Mok, P. P. K.: Rhythmic Correspondence between Music and Speech in English Vocal Music, *Proc. of Speech Prosody*, pp. 723–727 (2014).

[5] Patel, A. D. and Daniele, J. R.: An Empirical Comparison of Rhythm in Language and Music, *Cognition*, Vol. 87, pp. 35–45 (2003).

[6] Patel, A. D., Iversen, J. R. and Rosenberg, J. C.: Comparing the Rhythm and Melody of Speech and Music: The Case of British English and French, *Journal of Acoustic Society America*, Vol. 119, No. 5, pp. 3034–3047 (2006).

[7] Huron, D. and Ollen, J.: Agogic Contrast in French and English Themes: Further Support for Patel and Daniele, *Music Perception*, Vol. 21, No. 2, pp. 267–271 (2003).

[8] Sadakata, M., Desian, P., Honing, H., Patel, A. D. and Iversen, J. R.: A Cross-cultural Study of the Rhythm in English and Japanese Popular Music, *Proc. of International Symposium on Musical Acoustics*, pp. 41–44 (2004).

[9] VanHandel, L. and Song, T.: The Role of Meter in Compositional Style in 19th Century French and German Art Song, *Journal of New Music Research*, Vol. 39, No. 1, pp. 1–11 (2010).

[10] Temperley, N. and Temperley, D.: Music-language Correlations and the “Scotch Snap”, *Music Perception*, Vol. 29, No. 1, pp. 51–63 (2011).

[11] McGowan, R. W. and Levitt, A. G.: A Comparison of Rhythm in English Dialects and Music, *Music Perception*, Vol. 28, No. 3, pp. 307–313 (2011).

[12] 河瀬彰宏：“日本民謡の大規模音楽コーパスを用いた旋律の構造抽出”，国立国語研究所論集，No. 7, pp. 121–150 (2014).

[13] 持橋大地，山田武士，上田修功：ベイズ階層言語モデルによる教師なし形態素解析，情報処理学会研究報告 2009-NL-190 (2009).

[14] Teh, Y. W.: A Bayesian Interpretation of Interpolated Kneser-Ney, *NUS of Computing Technical Report TRA2/06* (2006).

[15] Stolcke, A.: SRILM – An Extensible Language Modeling Toolkit, *International Conference on Spoken Language Processing*, pp. 901–904 (2002).

[16] Neubig, G., 三村正人，森信介，河原達也：ベイズ推論を用いた連続音声からの言語モデル学習，情報処理学会研究報告 2010-SLP-82 (2010).

[17] Hsu, C.-W., Chang, C.-C. and Lin, C.-J.: “A Practical Guide to Support Vector Classification”，(online), available from <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (2010).